数据驱动下人文社会科学领域研究态势分析*
——基于 2010－2019 年国家社会科学项目的实证研究

翟姗姗[1] 叶丁菱[2] 许鑫[2,3]
[1] 华中师范大学信息管理学院 武汉 430079 [2] 华东师范大学经济与管理学部 上海 200062
[3] 华东师范大学调查与数据中心 上海 200241

摘 要：［目的/意义］在数据成为关键生产要素和基础战略资源的背景下，梳理人文社会科学领域中数据相关研究的发展现状，揭示数据问题的研究态势，为透视数据驱动下的人文社会科学研究提供窗口。［方法/过程］以 2010－2019 年间与数据有关的国家社会科学基金立项项目为研究对象，基于国家社会科学基金项目的外部特征与内容特征，分析人文社会科学领域中数据相关研究的机构及学科分布、关键词与主题聚类，对比国家社会科学基金项目和项目成果间研究热点的共同性与差异性，探测数据赋能的人文社会科学研究趋势。［结果/结论］人文社会科学领域中数据问题的研究以大数据环境、信息行为与信息服务、政府数据和科学数据研究为重点长线主题，以数据库建设及系统设计、媒体平台及服务和国际问题研究为次要热点主题，以融合多元环境、开放数据管理、延伸知识组织、智慧人文服务为研究发展趋势。

关键词：数据 人文社会科学 国家社会科学基金 研究特征
分类号：G250
DOI：10.13266/j.issn.0252－3116.2021.07.002

chinaXiv:202304.00659v1

数字经济背景下，数据的价值逐渐突显并成为社会发展的关键性动力。2020 年 4 月 9 日，我国出台的《关于构建更加完善的要素市场化配置体制机制的意见》将数据纳入生产要素范畴，提出加快培养数据要素市场的要求，充分体现了数据的基础性战略资源地位[1]。作为对社会生活与社会发展的学术表征，人文社会科学（以下简称"人文社科"）研究同样视数据资源为重要生产要素和核心创新资源，以数据为驱动促进人文社科研究的纵深化发展。一方面，数据驱动模式下新视角、新热点的产生，激发人文社科的研究活力，拓展人文社科研究的外延空间；另一方面，在数据密集型科学研究范式下新特征、新路径的形成，冲击人文社科的传统认知，延伸人文社科研究的内涵边界。为在新范式、新场景下把握人文社科领域中与数据有关研究的发展情况与动态趋势，本研究以 2010－2019 年这 10 年间国家社会科学基金（以下简称"社科基金"）各类型立项项目为样本，利用数据统计与可视化工具进行分析，解析近 10 年来人文社科领域中数据研究的发展态势，理清数据有关问题的发展格局、脉络与前沿，为透视数据驱动下的人文社科研究提供窗口以及为新文科的建设提供新视角。

1 数据驱动的人文社科领域研究现状

数据作为人文社科研究人员认识研究问题的新逻辑起点，突破了基于文献和知识探索人文社科研究的思维桎梏，引发众多学者对人文社科领域中数据问题的不同视角的探讨。

在基础理论研究方面，相关学者主要考量数据思维的引入为人文社科研究带来的价值与风险。倪万等针对数据应用于人文社科研究中的基础问题，指出两者的结合具有原则使用总体数据实则使用样本数据、数据时空混杂影响精准性、数据反映相关性缺乏因果性等问题，因而人文社科研究对于数据思维的引入要充分考量自身研究特性与数据的作用方式[2]。陈泓茹

* 本文系国家社会科学基金项目"图书馆智慧空间的理论构建与实践应用研究"（项目编号：18BTQ028）研究成果之一。
作者简介：翟姗姗（ORCID：0000－0002－2787－0183），副教授，博士；叶丁菱（ORCID：0000－0002－1306－2919），博士研究生；许鑫（ORCID：0000－0001－7020－3135），调查与数据中心副主任，教授，博士，博士生导师，通讯作者，E-mail：xxu@infor.ecnu.edu.cn。
收稿日期：2020－10－26 修回日期：2021－01－12 本文起止页码：15－24 本文责任编辑：徐健

ChinaXiv合作期刊

等指出数据思维与人文社科研究的融合需要确定基本的作用限度：其一，数据改变人文社科研究的方式与存在形态，但无法促使研究人员精神世界数据化；其二，充分揭示与认识数据具备必要性，但需要警惕对数据的崇拜行为；其三，合理应用数据的同时需要关注数据引发的隐私、安全等伦理问题；其四，数据可以提升人文社科研究的精准程度，然而精准化并不等同于研究的科学化[3]。是沁等指出人文社科研究过程中存在数据需求迫切与供给不足、数据安全管护与隐私侵犯风险、数据客观真实与篡改造假等矛盾，人文社科研究需要促进数据资源的开放、构建数据管理和监督机制、加强数据素养的培训和教育[4]。

在关键技术应用方面，相关学者主要探究数据相关技术或平台为人文社科研究场景和组织架构带来的变革。马费成指出在数据驱动人文社科的新场景下，人文社科的研究需要开发与促进数据获取、存储、分析与呈现阶段的有形研究工具，通过新的技术工具与思路方法改进和创新人文社科研究[5]。王晓光认为人文社科领域的发展必须以数据资源建设为驱动，以专题数据库与计算平台为保障，通过数据资源评价标准的建立、数据库建设流程的规范、数据模型技术的研发、数据交汇分析工具的运用，管理与建设人文社科数据资源[6]。谷俊等针对人文社科中数据溯源能力弱、数据用途追踪难等问题，采用区块链技术设计并构建了人文社科数据共享联盟平台，实现数据利用的共享、溯源与追踪[7]。同时，人文社科数据开发的应用实践也在众多高校或研究机构开展，如复旦大学社会科学数据平台旨在获取、组织和开发社会科学数据，以为研究人员提供丰富、及时的数据服务[8]；南京大学人文社会科学大数据研究院旨在打破人文社科的信息壁垒，整合、发展和提供数据资源与技术，以推动人文社科的研究发展与学科建设[9]。

在服务体系构建方面，相关学者主要从服务层面探讨数据赋能模式下人文社科领域的研究创新和智慧服务模式的建设。李阳等通过对人文社科数据研究的缘起、脉络和特征的解析，指出人文社科数据研究需要从数据展示的浅层化研究向知识标识的智慧化和意蕴化方向发展，实现具有中国特色的人文社科数据研究体系[10]。米加宁等认为数据驱动的范式推进了人文社科研究的认识论与方法论，变革了人文社科研究的思路与模式，构建了多元集成的人文社科学术研究体系，但同时要注重人文社科研究中数据应用技术的伦理问题和人文社科新场景下的价值挖掘问题[11]。刘雨农等通过分析数据驱动模式下人文社科研究的现实困境和诉求，指出人文社科创新研究与智慧服务的构建一方面需要为研究人员开展数据定制、协助处理、技术指导、智慧推送等服务，另一方面需要具备数据安全可靠、服务专员培训、数据产权明晰等制度保障[12]。

综合来看，学者们对于数据驱动人文社科领域的研究主要体现在3个方向：①在泛在知识环境下，考量引入数据思维和数据背景对人文社科研究带来的理念变革；②关注数据相关技术或方法对人文社科研究带来的发展机遇以及对人文社科领域中各类社会现象的解读和知识发现平台的建设；③注重人文社科领域的数据生态体系构建，以系统化的服务促进人文社科中数据研究的智慧化发展。3个研究方向层层递进，在一定程度上描述与揭示了数据融于人文社科研究中的共性问题。在现有学者研究的基础上，为把握数据应用于人文社科的研究动态、研究热点以及研究前沿，本文基于文献计量，在整体上对社科基金立项项目中数据有关项目进行多视角的实证研究，在局部分析社科基金项目成果论文的知识结构，以从宏观上把握数据驱动的人文社科研究的发展态势，从微观上揭示数据有关研究的细粒度发展特征。

2 研究设计

2.1 研究思路

社科基金作为我国在人文社科领域最具权威与影响的项目资助和管理组织，其立项项目通常侧重于理论前沿问题、政策导向问题和重大发展问题的研究[13]，在一定程度上反映出人文社科研究的前瞻性、发展性与交叉性，折射出人文社科研究的动态趋势与发展态势。因此，本研究首先对社科基金项目中与数据相关项目的年度立项数量与项目类别、工作单位与学科类别进行统计分析，从立项结构纵观数据赋能的人文社科研究的外部特征与整体发展趋势。随后对社科基金项目及项目成果进行关键词分析与研究热点对比分析，关键词作为对研究内容的精确凝练，能在一定程度上代表研究的重点与热点。研究热点对比分析则通过预期研究与实际研究的内在内容比较，在较大程度上实现人文社科领域中数据研究的发展特征解析。最后综合立项项目的研究主题与知识特征，探测与发掘数据驱动范式下人文社科领域的研究生长点。

2.2 数据来源

本研究样本数据来自全国哲学社会科学工作办公室网站中"项目查询"下设的"立项查询"系统[14]，以

chinaXiv:202304.00659v1

chinaXiv:202304.00659v1

“数据”作为关键词进行搜索，搜集与统计2010－2019年间与数据相关的各项立项项目，获得有效立项项目952项，具体获取项目批准号、项目类别、学科分类、项目名称、立项时间、项目负责人、专业职务、工作单位、单位类别等字段内容。同时，使用中国知网获取基金立项项目资助的科研论文的题录信息，共获取科研论文5 293篇。

2.3 研究方法

本研究主要采用文献计量法、聚类分析法和内容分析法。使用人工判读的形式，统一立项项目单位名称，如将“中山大学图书馆”统一为“中山大学”。使用Excel与SPSS 25.0对研究数据进行汇总、统计与图表绘制，实现对研究数据的外部特征分析。通过ROST CM6.0文本内容挖掘平台实现对立项项目和科研论文的分词处理，同时通过人工识别的形式，对关键词进行筛选与合并，如剔除“研究”“模式”“分析”等无特别区分度的词语，合并“中国”和“我国”等表述不同但内涵相同的词语。通过Ucinet 6.186软件实现关键词共现矩阵，使用Netdraw 2.084可视化工具生成关键词共现图谱，利用Gephi 0.9.2对研究数据进行内容挖掘处理，实现主题聚类分析。最终在此基础上，实现对研究热点、研究趋势的内容特征分析。

3 人文社科的数据研究支撑与特点分析

3.1 立项年度分布分析

社科基金项目年度立项数量的变化可以在一定程度上反映学术研究的发展水平和演化过程，也可以从侧面体现学术研究的重要程度。2010－2019年社科基金项目中数据相关项目的立项情况如图1所示：

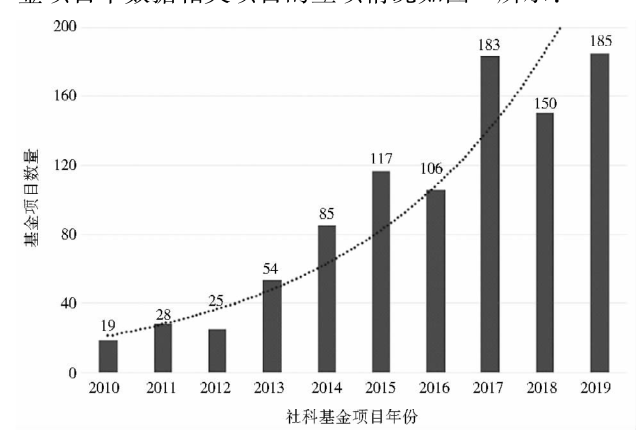


图1 数据相关社科基金项目年度统计

从图1可知，社科基金项目中数据问题的相关研究开始于2010年，兴起于2013年。这与2009年数据密集型科学研究第四范式的提出、2013年大数据技术的广泛应用具有直接相关性，说明人文社科中数据问题的研究与社会发展、学术环境、信息技术具有紧密关联。对各年度的社科基金项目进行趋势拟合，发现社科基金项目年度数量较符合指数型分布，前期增长缓慢，后期连续6年立项数量达100项以上，说明数据问题的相关研究受到国家与研究人员的持续关注，重视程度与发展程度呈现稳步提升的趋势，该研究具有较好的发展空间与生长潜力。

为掌握数据相关立项项目的具体分布情况，明确数据相关问题的研究粒度，笔者对社科基金项目按立项时间与项目类别进行拆分统计。社科基金项目按照资助额度、学者年龄、地区属性的不同，共设有8类立项级别，数据相关基金项目的年度项目类别及其数量分布情况如表1所示：

表1 数据相关社科基金项目类别分布

年份	重大项目	重点项目	一般项目	青年项目	西部项目	后期资助项目	中华学术外译项目	合计
2010	3	1	8	4	2	1	0	19
2011	4	3	12	6	3	0	0	28
2012	6	2	7	8	2	0	0	25
2013	9	10	14	15	5	1	0	54
2014	13	13	38	19	2	0	0	85
2015	22	7	56	25	5	2	0	117
2016	0	5	75	19	7	0	0	106
2017	39	16	81	33	11	2	1	183
2018	0	8	96	28	18	0	0	150
2019	33	12	96	35	9	0	0	185
合计	129	77	483	192	64	6	1	952

在所有项目类别中，一般项目占据主体地位，青年项目次之，在一定程度上说明国家对中青年学者的鼓励与注重，中青年学者成为人文社科中数据问题研究的中坚力量，预示该研究具备良好的发展前景。重大项目把控严格、要求较高、立项困难，数据相关社科基金项目中重大项目数量较多，一方面反映出国家对数据问题的相关研究重视程度较高，数据问题的研究与社会需求契合程度较强；另一方面显示出人文社科视域下数据问题的研究具有较大影响力，研究深度正在逐步加强，研究体系正在逐步成熟。

3.2 研究机构分析

分析社科基金项目中工作单位的分布，可以了解数据相关研究的分布情况，从侧面反映出在研究中具备较高科研水平的机构。在952项基金项目中共有316个机构负责立项项目，平均每个机构负责3个项

ChinaXiv合作期刊

目。然而负责 3 个及以上项目的机构仅 96 个,说明人文社科视域下数据问题的研究呈现出集中与分散并存的不均衡现象。对项目数量在 10 个及以上的工作单位进行统计分析,具体如表 2 所示:

表 2　数据相关社科基金项目工作单位统计

工作单位	项目数量	所属地域
中国人民大学	21	北京
中山大学	15	广东
南京大学	15	江苏
清华大学	14	北京
上海交通大学	14	上海
华中师范大学	11	湖北
复旦大学	11	上海
北京师范大学	11	北京
武汉大学	11	湖北
厦门大学	10	福建
华南理工大学	10	广东
中央民族大学	10	北京

3.3　学科分布分析

社科基金立项中数据相关项目的学科分布特征是相关学理支撑与方法支持交互作用的产物,分析学科演变特点有助于把握人文社科领域中数据研究的学科建设情况,探测数据研究的外部发展方向与动态趋势。为直观分析数据相关研究的学科发展变化,绘制如图 2 所示的学科演变图谱。

从学科分布演变情况来看,在整体上,人文社科领域中与数据相关的研究呈现百花齐放的样态;在局部上,图书馆、情报与文献学、统计学和管理学 3 个学科的社科基金项目数量与项目增长量明显突出,是数据研究的主要学科。新闻学与传播学、语言学两个学科的社科基金项目数量居中,呈现持续增长状态,是数据研究的潜在发展学科。

图书馆、情报与文献学作为对数据资源进行获取、组织和利用,从数据中提炼和挖掘信息, 形成知识理

chinaXiv:202304.00659v1

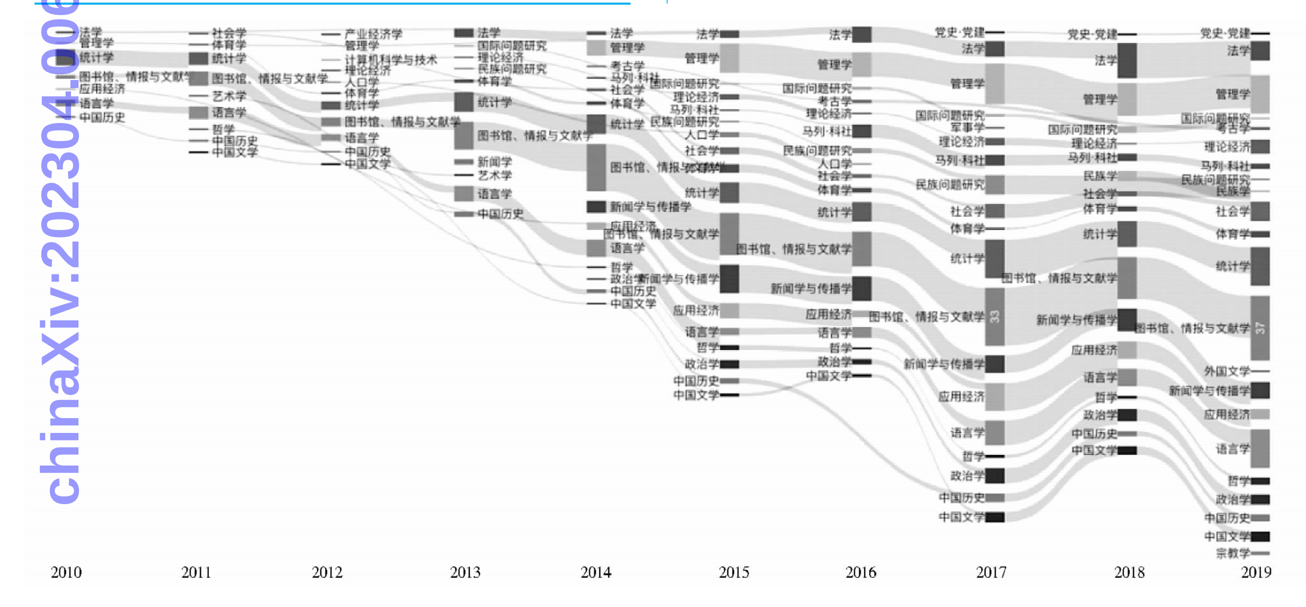


图 2　数据相关社科基金项目学科分布演变图谱

论、技术方法、应用工具的学科[15],自身就是数据资源的重要集散地,因而展开数据问题的研究是图书馆、情报与文献学发展的必然之势和长久之道。统计学作为植根于数学的学科,数据贯穿于学科研究的始终,因此对数据问题的研究必然会较为突出。管理学是对社会管理活动进行研究的一门学科,具备广泛性与交叉性[16],数据思维和技术方法的发展,逐渐改变社会的生产力与生产关系,引发管理要素、管理思维、管理边界的持久变化,促使管理学科的研究转向数据要素带来的长期管理变革。因此,图书馆、情报与文献学、统计学和管理学 3 个学科既是人文社科视域下数据问题现有研究的主要学科,也是未来深度发展的重要支撑学科。

新闻学与传播学作为对新闻活动的传播以及新闻与社会关系研究的学科,本质是对数据加工形成的有效文本信息进行研究[17],其学科的发展有赖于对数据资源的挖掘、利用与服务。语言学是对语言的本质、发展规律进行研究的一门学科[18],需要对大量语料数据进行组织分析,得出统计规律,实现对语言系统运作机理的解释和对语言结构演化规律的认知。作为对数据资源加工形成的二次信息资源的分析和对数据技术方法的应用,新闻学与传播学、语言学的建设与研究发展

ChinaXiv合作期刊
chinaXiv:202304.00659v1

需要长期在数据驱动模式下拓展其研究要素、深化其学科内涵、创新其学科体系。因此,新闻学与传播学、语言学的发展现状与长期发展要求预示其将成为人文社科视域下数据问题研究的主要趋势学科。

4 人文社科的数据研究重点分析

分析数据问题的研究特点可以从宏观层面展现人文社科领域中数据研究的外部演化特征,为深入探析数据问题的具体研究重点,还需从微观层面揭示数据研究的内在内容特征。为全面发掘数据问题的研究重点,透视数据研究的发展趋势,本研究对社科基金项目及其资助的科研论文分别进行研究热点识别,并展开对比分析。

4.1 立项项目关键词及主题内容分析

本研究的数据源为社科基金项目名称,项目名称一般由研究角度、背景和内容组成,可以从侧面体现立项项目的核心要点[19]。通过对立项项目进行关键词分析,生成社科基金项目关键词图谱,见图 3(Top20,频次≥43)。借鉴大数据模型中的降维和聚类算法,通过模块度和随机算法进行主题聚类分析,梳理并归纳社科基金立项项目的研究热点和方向,如图 4 所示:

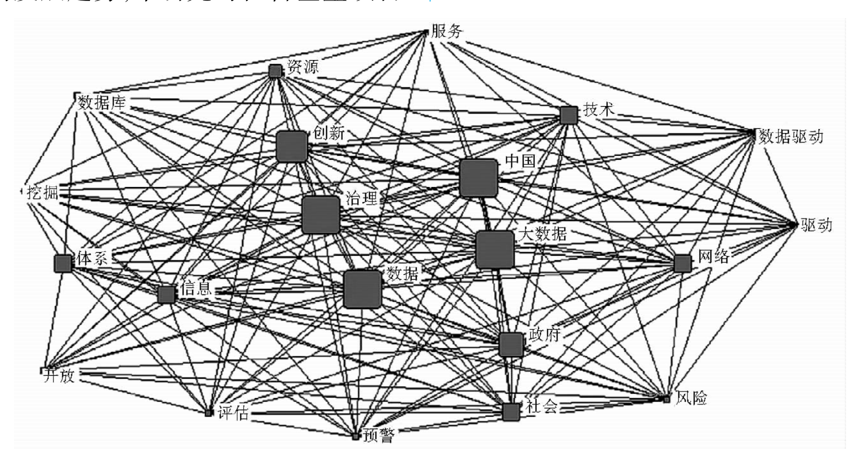


图 3 数据相关社科基金项目关键词图谱

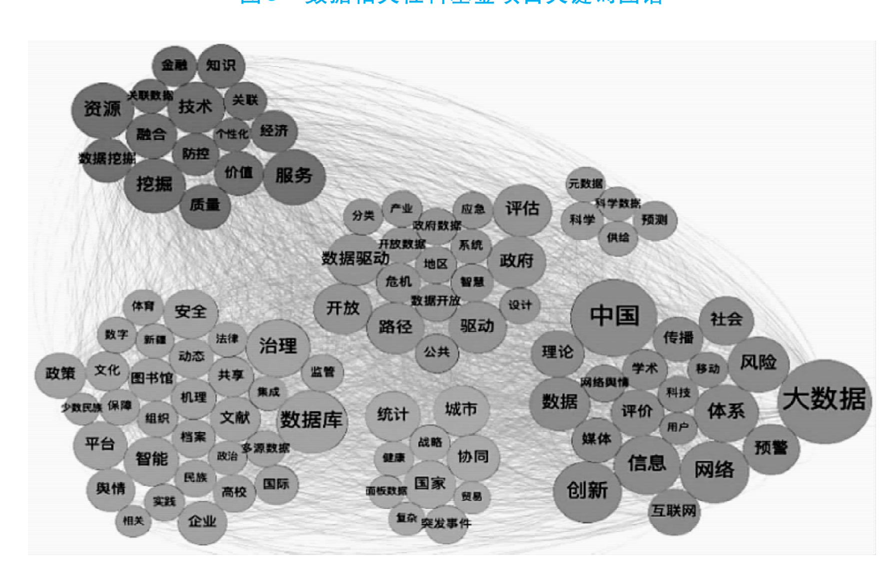


图 4 数据相关社科基金项目研究热点

从图 3 可以看出,人文社科视域下数据问题的研究主要以大数据、数据库、中国、治理、创新等为主要关键词。“大数据”出现的频次为 492 次,具有绝对优势。在 2013 年大数据元年后,大数据以势不可挡的态势在各行各业探索实践,改变各领域的生产经营模式和管理形式,影响大众的思想意识和生活方式,因而大数据不可避免地成为人文社科领域中数据问题的关键研究要点。“数据库”出现 206 次,说明对资源进行组织、建设和保存仍是研究人员较为关注的问题。“中国”出现 123 次,这与国家社科基金项目是从国家层面出发,统筹考虑和研究我国社会发展问题具有直接相关性。“治理”出现 62 次,这与十八届三中全会提出从社会管

ChinaXiv合作期刊

理转向社会治理有关[20]，说明人文社科的研究紧跟国家发展方略和导向，注重对社会突出问题的解决，体现人文社科研究的现实意义。“创新”出现 58 次，说明数据驱动的新发展方式和思维方法在促进社会进步、创新发展的同时，为人文社科研究也带来新的发展路径或模式。

为更加精准地分析数据问题的研究重点，本研究通过 Gephi 团体发现分析进行聚类，Gephi 利用随机算法、分割模块等实现群组分解与聚合，相对于 Ucinet 可以更加直接、优化地分解聚类结果，聚类结果一般通过节点大小和节点颜色进行标识，节点越大则表示关键词的关联影响力越强[21]，具体见图 4。通过人工判读，发现突发事件相关立项项目皆以大数据为背景，可以与大数据环境研究的类目合并，即近 10 年来数据相关的基金项目研究可以总结为：以大数据环境为重要研究背景，注重信息服务、数据开放等服务行为的研究，注重数据库建设等保障行为的研究。反映出人文社科领域注重将社会发展的时代背景与科学研究有机结合，体现人文社科研究的社会价值，同时，注重用户、政府和研究人员的数据服务行为，体现人文社科研究的人文价值。

（1）基于大数据环境的研究。作为当今时代发展的重要背景，大数据既是社会发展的一种现象，也是人文社科领域中数据研究的绝对热点。通过对项目主题的深入分析发现，大数据与人文社科领域充分融合，利用大数据思维能够全面深化研究内容，解决社会发展问题，构建学科发展体系，突出人文社科的研究价值。一方面，大数据时代引发全面变革，从国家战略发展到社会预警治理，从社会管控应急到网络信息安全、从资源环境优化到产业结构升级、从学科发展创新到教育方式变革，社会发展、管理模式、思维方式、科学研究等各方面已经与大数据紧密结合。另一方面，大数据作为技术方法，以其多元海量、层级复杂、全面真实的计算特征，冲击人文社科领域的传统研究方法，催生出数据挖掘、机器学习、聚合计算等新兴技术与方法，促进人文社科研究的全领域、深层次和创新化发展。细化而言，通过对 2013 – 2019 年图书馆、情报与文献学有关大数据立项项目的分析可以发现，大数据立项项目大致可以分为 3 个阶段，第一阶段是大数据背景下，有关信息安全、图书馆用户服务等的研究；第二阶段是大数据技术下，有关多层级文本挖掘、语义挖掘、舆情监测、企业服务等的研究；第三阶段是大数据生态下，有关智库、人才库构建与评价、科教评价平台构建、档案服务平台构建等的研究。3 个阶段的大数据立项项目也与前述诸多学者对数据驱动人文社科领域的 3 个研究方向相契合。

（2）信息行为与信息服务研究。满足用户信息需求的服务研究也是人文社科领域中数据研究的重点主题，反映出对民生问题的高度重视和对实际问题的有效解决能力。运用多源数据和技术方法，将受众为主要形式的被动服务转化为以用户为中心的主动服务模式，进而通过个性化推荐和精准供给满足用户各层次的需求是信息服务研究的主要内容。在信息服务对象方面，包括学生、青年、老年人和企业，用户群体多样、层次分明，能切实有效地反映或解决各层次用户的不同信息需求服务；在信息服务主体方面，包括政府部门的公共服务、图书馆的智慧服务、档案馆的数字服务，注重公共服务体系对用户信息需求的预测、用户行为的掌握、服务能力的提升和服务质量的改进研究；在信息服务内容方面，关注基于知识的智慧信息服务、基于本体的个性化推荐服务、基于感知的细粒度人性化服务等内容，优化用户服务体系，紧跟社会发展重点，体现研究的理论价值与实践意义。

（3）数据库建设及系统设计研究。数据库建设将分散无序的资源进行系统化、深层化和有序化组织，将传统资源抽取、整理和加工形成数字资源，利于网络传播、存取和使用，重现和拓展资源的使用价值。数据库建设的相关研究侧重以特色数据库为导向，实现资源的共享，拓宽服务的空间。研究内容呈现多元化特征，具体包含从初始的多民族语言词汇、非物质文化遗产、古籍文献资源等静态数据库的建设，到国际关系、用户行为、各国华文教育等动态数据库的建立，再到语音、影像等有声数据库的构建。说明数据技术的发展为资源提供了有效地开发、采集、著录、检索、交互等能力，同时各学科领域的特色数据库建设为人文社科研究的纵深化发展奠定了坚实的知识型数据基础。

（4）政府数据相关问题研究。政府数据蕴含丰富的社会价值与经济价值，激活数据资源可以充分释放政府数据的红利。政府数据相关的研究主要体现在 3 个方面，一是政府数据的开放研究，包括数据开放的体系建立、开放平台的机制建设、开放共享的风险防范和保护策略研究；二是政府数据的治理研究，包括治理机理的探究、社会舆情的治理策略、治理平台和体系的建立；三是政府数据的评价研究，包括廉政建设的评价、网络传播力的评价。说明政府数据以开放、融通、治理、应用和评价为关键，研究视角多元，研究内容多样。

chinaXiv:202304.00659v1

ChinaXiv合作期刊

同时,重视政府数据的应用效果,表明相关研究已从注重机理分析进入到关注实践评价的新阶段,反映出政府数据研究的系统性、务实性和实用性。

(5)科学数据相关问题研究。科学数据作为科学假设、科学分析以及科学理论形成的基础,较大程度上决定了科学研究的质量,并且科学数据的可再利用性、可再分析性和可信度使其具有可挖掘的扩展价值。科学数据的研究以发现、获取、理解和重用为关键,具体包括科学数据的开放政策研究、开放模式分析、关联关系挖掘、引用机制实现和影响因素探究,尤其是开放获取成为研究的主要焦点,表明科学数据的研究尚处于初期发展阶段,可以预见对其理论基础的深入挖掘、理论体系的构建、管理机制的探索、创新实践的应用将成为后续研究的重点。

4.2 项目成果关键词及主题内容分析

作为社科基金项目的重要承载形式,科研论文从选题确立到撰写发表的全过程都会经过研究人员的广泛研讨和深入论证,在较大程度上表征了立项项目的深度研究内容和具体研究方向,分析科研论文的关键词和主题词可以更具化地体现科研人员实际研究重点,以此窥探与拓展人文社科中的数据研究热点。由于立项项目的分析对象为标题,基于此,本研究获取立项项目资助的科研论文标题数据,通过论文标题分词、词频统计、关键词共现和聚类分析的形式,探测和分析项目成果的研究热点与重点。

从图5可以看到,科研论文的研究重点与立项项目的研究重点具有一定相似性,如“大数据”和“中国”具有绝对优势,但同时也发展出新的研究要点,如“网络”和“信息”,进一步验证了科研论文是对立项项目的具体表征,也是对立项项目的内涵发展。“中国”出现625次,“大数据”出现500次,“服务”出现187次,“开放”出现141次,说明基于大数据环境的研究、针对用户服务的研究和促进数据开放的研究仍然是数据相关研究的热点和前沿内容。“网络”出现326次,“信息”出现253次,“知识”和“技术”出现136次,信息是基础,知识是延伸,技术是手段,网络是平台,研究之间相互关联,呈现出系统化、泛在化和应用化特征。

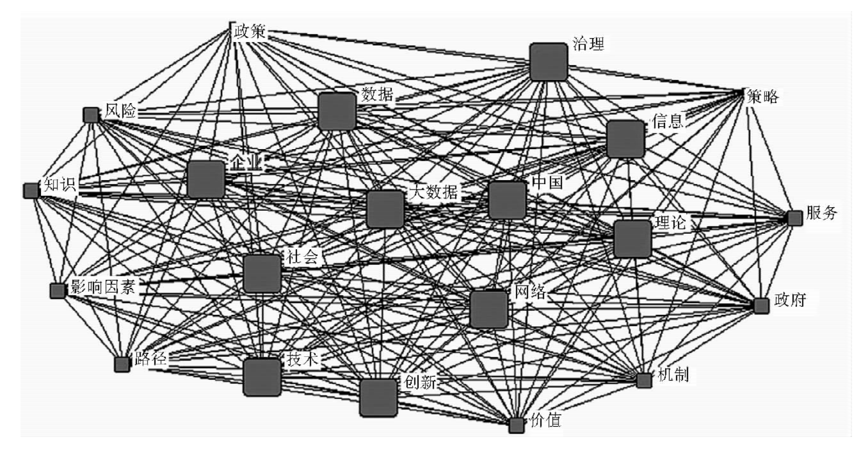


图5　数据相关科研论文的关键词图谱

根据图6所示的Gephi聚类结果,结合人工判读,可以看到有关大数据、信息服务和数据开放的研究依旧是数据研究的前沿主题,同时知识管理、新媒体平台和国际问题是数据研究的新动向,说明人文社科领域对数据问题的研究在实践中得以拓展,注重知识转化的智慧服务、社交媒体中的泛在信息传播和国际视野下的国家发展问题,紧跟社会发展热点,突显时代特征。

(1)知识组织与管理研究。知识作为创新的内在驱动,是促进社会发展、激发科研活力的重要推动力。人文社科领域中数据问题的研究呈现出“数据－信息－知识”的发展脉络,从数据开发、获取等表象行为为深入到信息资源的建设与管理过程,再到语义分析、知识生产、知识挖掘、知识共享等知识的组织和服务模式,说明数据问题的研究从信息化逐渐转向知识的发掘利用和知识需求的满足服务,知识化成为数据研究的重要内容。

(2)媒体平台及服务研究。社会媒体是以互联网技术为基础,以用户生成信息为内容,以多向传播为目的的大众化媒体,包含丰富的多源信息,蕴含巨大的社会价值、经济价值与学术价值。社会媒体的相关研究主要体现在两个方面,一是在泛在信息环境下,利用用户信息数据以及大数据技术实现用户行为与信息传播模式的研究;二是在信息异化的现象下,展开用户隐私

chinaXiv:202304.00059v1

ChinaXiv合作期刊

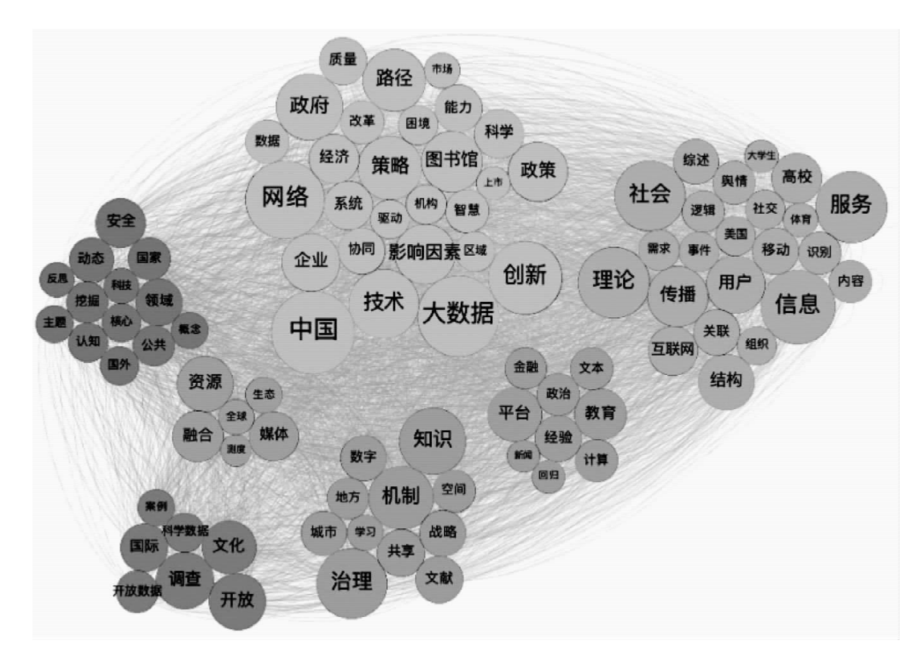


图 6　数据相关科研论文的研究趋势

保护与动态监测管理的研究。说明数据研究在注重挖掘社会媒体数据资源的同时,关注人性化发展与人文关怀。

(3)国际问题相关研究。利用数据资源与数据技术,获得全面的事实信息是国际问题研究的新方式。国际问题的研究主要关注国际关系问题与国际经济问题,具体包括“一带一路”倡议、国际竞争力分析、国际公共卫生合作、国际关系冲突预测、资本市场竞争、国际资本流动、国际贸易关系等。在无边界网络环境下,各国交往日益密切,各要素流动自由便利,如何把握数据资源创造新的发展机遇,提升能级是人文社科领域中数据研究必须重视的问题。

4.3　研究热点对比分析

社科基金项目作为社会发展问题的表征具有全局性和概括性,项目成果作为社科基金项目在实际研究中的高度凝练,具有延展性和粒度化。通过对二者的研究热点进行对比与融合,既可体现数据驱动人文社科的预期研究前沿,又可表征实践研究中的实际内容应用,从而在较大程度上解析人文社科领域中数据研究的特征与价值。可以看到,大数据背景、信息服务和数据开放 3 个研究主题是数据研究中的重要长线主题,数据库建设、新媒体平台和国际问题 3 个研究主题是数据研究的次要热点主题。总体而言,社科基金项目和项目成果的研究热点呈现关键主题趋同,部分主题存异的特点。

4.3.1　关键研究主题趋同

大数据背景、信息服务和数据开放作为长线研究主题在社科基金项目和项目成果中较受注重,并且均与大数据发展战略、信息服务战略、数据开放纲要等国家重大战略思想和国家宏观政策相契合。同时注重服务于“人”的研究,如大数据背景下的消费者行为研究、个人信息安全研究、科研数据用户使用模式研究等,体现出数据研究注重人文关怀和人本服务。人作为社会的主体,是数据资源的使用者、知识文化的创造者、创新发展的推动者,满足用户需求、便利用户行径、保障用户权益是人文社科研究的经久重点和必然要点。

4.3.2　部分研究主题存异

社科基金项目注重对数据库的建设研究,说明社科基金项目重视数据资源的基础性和支撑性作用。专题数据库的建设通过对数据的处理和组织,实现数据的集合,便于数据资源的控制、调取、维护和共享,充分发挥数据的可用价值,为人文社科研究在数字空间中的繁荣发展奠定资源基础。项目成果则注重对新媒体平台与国际问题的研究,说明项目成果重视社会现象,时效性、贴合性强。自大数据等先进技术兴起、互联网+新业态的发展,新媒体平台日趋兴盛,越发成为用户生活中不可或缺的一部分,新信息环境下对用户行为模式的新特征探索一直作为社会交互和科学研究的热点。在国际联系越发紧密,我国相继提出“加快培育国际合作和竞争新优势指导意见”、共建“一带一路”的顶层合作倡议等政策后,我国已置身于国际视野,在不同社会发展背景和国际关系中探寻国际问题,发掘竞争优势成为国家发展、社会研究的持续热点内容。

chinaXiv:202304.00659v1

ChinaXiv合作期刊

5 人文社科的数据研究趋势透视

梳理国内有关人文社科领域中数据研究的文献内容，结合数据问题的研究热点，本研究认为未来人文社科视域下数据问题的研究将在研究背景、研究对象、研究内容和研究体系中实现突破，呈现情景综合化、应用广泛化、资源知识化和服务人本化的特征趋势。

融合时代特征，紧跟时代潮流。人文社科研究深受国家战略、政策制度、社会环境和国际关系的影响，必须紧抓时代发展潮流，保持研究的有效性和先进性。在5G时代到来、大数据技术深化、国际竞合紧密的背景下，数据实现泛在聚合与共享，数据的协同价值越发突出，如何在新时代背景下，充分实现数据驱动的热点研究内容持续化和深度化发展，融入新概念与新视角以挖掘研究生长点，保持研究的增长活力是人文社科领域中数据研究必须持久把握的问题。

融入数据控制，开放数据管理。开放数据一方面实现数据的流动与共享，另一方面改变知识交流的生态，推动经济社会的发展[22]。开放数据研究是数据相关研究的重要组成部分，然而有关研究尚处于探索发展阶段，具有较大发展前景。如何增强数据共享意识、制定开放标准、保障开放机制、完善基础设施、促进实践应用和建立评价体系都是开放数据研究中亟待解决的问题。

重视数据挖掘，延伸知识组织。数据思维和数据技术促使信息组织内容发生变化，从数字化阶段向数据化变革。信息组织工具从传统受控词表向语义网络发展，信息组织内容则基于语义层级的变化向知识组织系统转变。在未来，如何通过语义组织、语义分析、知识挖掘和知识识别集成元数据、语义数据和关联数据是实现语义化知识组织和管理的有效突破口。

数据作为资源，智慧人文服务。科学研究价值的实现需要服务于人类与社会，注重用户需求、关注用户行为、提升用户服务，是人文社科研究的价值所在。在数据驱动范式下，智慧数据发展、智能技术升级，实现智慧与人文的耦合，实现数据成果与人文服务的同频共振，是人文社科领域中数据研究的重要导向。

6 结语

本研究对近10年的数据相关社科基金项目展开分析，旨在把握人文社科研究中与数据有关的研究支撑与特点以及研究重点与趋势，形成对数据研究的整体性认识。在研究支撑与特点上，数据研究存在学科与地域偏差。为实现数据研究的可持续发展，今后应注重跨学科、跨地域的有机结合与协同合作，促进数据研究的横向渗透交叉和纵向继承发展。在研究重点与趋势上，数据研究以大数据与国际环境为背景，形成“数据－信息－知识－服务”的研究主线，研究层次丰富饱满。数据为人文社科研究带来生机与活力，协同数据理论的本源思考与数据应用的本位研究，促进人文社科研究的智慧化转型是人文社科领域中数据研究的重要态势。

参考文献：

[1] 中华人民共和国中央人民政府.《关于构建更加完善的要素市场化配置体制机制的意见》[EB/OL].[2021－01－03]. http://www.gov.cn/zhengce/2020-04/10/content_5500740.htm.

[2] 倪万，唐锡光. 大数据应用于社会科学研究的价值与悖论[J]. 东南学术，2017(4):68－78,247.

[3] 陈泓茹，赵宁，汪伟. 大数据融入人文社会科学的基本问题[J]. 社会科学文摘，2016(2):16－18.

[4] 是沁，李阳. 数据驱动的人文社科研究困境与对策:基于微观数据层面的考量[J]. 图书与情报，2019(1):17－23.

[5] 马费成. 推进大数据、人工智能等信息技术与人文社会科学研究深度融合[J]. 评价与管理，2018,16(2):1－5.

[6] 王晓光. 加强人文社科数据资源建设与管理[EB/OL].[2021－01－03]. http://theory.people.com.cn/n1/2018/0705/c40531-30127206.html.

[7] 谷俊，许鑫. 人文社科数据共享模型的设计与实现——以联盟链技术为例[J]. 情报学报，2019,38(4):354－367.

[8] 复旦大学社会科学数据平台[EB/OL].[2021－03－29]. http://fisr.fudan.edu.cn.

[9] 南京大学人文社会科学大数据研究院[EB/OL].[2021－01－03]. https://skbigdata.nju.edu.cn/data/.

[10] 李阳，孙建军. 人文社科大数据研究的价值追寻[J]. 图书与情报，2019(1):1－7.

[11] 米加宁，章昌平，李大宇，等. 第四研究范式:大数据驱动的社会科学研究转型[J]. 学海，2018(2):11－27.

[12] 刘雨农，是沁. 数据驱动范式下的人文社科知识服务创新研究[J]. 图书与情报，2019(1):24－30.

[13] 巫程成，朱倩倩. 我国文旅融合研究学术态势——基于2009－2019年国家社科、教育部人文社科基金立项数据[J]. 旅游论坛，2019,12(6):64－74.

[14] 全国哲学社会科学工作办公室[EB/OL].[2021－01－03]. http://www.nopss.gov.cn.

[15] 苏新宁. 大数据时代情报学与情报工作的回归[J]. 情报学报，2017,36(4):331－337.

[16] 陈劲，尹西明. 范式跃迁视角下第四代管理学的兴起、特征与使命[J]. 管理学报，2019,16(1):1－8.

[17] 骆正林. 新闻学的学科内涵及新闻学理论体系的创新[J]. 新闻界，2019(4):11－26.

[18] 梁君英，刘海涛. 语言学的交叉学科研究:语言普遍性、人类认知、大数据[J]. 浙江大学学报(人文社会科学版)，2016,46

chinaXiv:202304.00559v1

ChinaXiv合作期刊

(1):108-118.
[19] 廖嘉琦. 我国情报学近五年研究热点及发展趋势分析——基于2014-2018年国家社科基金立项[J]. 情报科学,2020,38(3):160-166.
[20] 中国共产党第十八届中央委员会第三次全体会议公报[EB/OL]. [2021-01-03]. http://www.xinhuanet.com//politics/2013-11/12/c_118113455.htm.
[21] 邓君,马晓君,毕强. 社会网络分析工具 Ucinet 和 Gephi 的比较研究[J]. 情报理论与实践,2014,37(8):133-138.
[22] 吴建中. 推进开放数据 助力开放科学[J]. 图书馆杂志,2018,37(2):4-10.

作者贡献说明:

翟姗姗:提出论文研究思路与框架;
叶丁菱:采集与分析数据,撰写初稿;
许鑫:提出论文修改意见并修订最终版。

Analysis on the Research Situation of Humanities and Social Sciences Field Driven by Data ——Empirical Research Based on 2010-2019 National Social Science Project

Zhai Shanshan[1] Ye Dingling[2] Xu Xin[2,3]
[1] School of Information Management, Central China Normal University, Wuhan 430079
[2] Faculty of Economics and Management, East China Normal University, Shanghai 200062
[3] Social Survey and Data Center, East China Normal University, Shanghai 200241

Abstract: [Purpose/significance] Under the background that data has become a key factor of production and basic strategic resources, this paper reviewed the research status of data data-related research in the field of humanities and social sciences, reveals the research trend of data problems, and provided a window for the perspective of humanities and social sciences research driven by data. [Method/process] Taking the National Social Science Fund Project about "data" of 2010-2019 as the research object, this paper based on the external characteristics and content characteristics of the fund project, analyzed the distribution of data-related research institutions and disciplines, keyword and topic clustering in the field of humanities and social sciences, compared the commonness and differences of research hotspots between fund projects and project achievements, and explored the research trends of humanities and social sciences empowered by data. [Result/conclusion] The research on data issues in the field of Humanities and Social Sciences focuses on the long-term theme of big data environment, information behavior and information service, government data and scientific data research, the secondary hot topics are database construction and system design, media platform and service and international issues research, and the integration of multiple environment, open data management, extension of knowledge organization and wisdom humanities service are the research development trend.

Keywords: data humanities and social sciences the National Social Science Fund research characteristics

chinaXiv:202304.00659v1